THE GENERALIZED COMBINATORIAL
LASOÑ-ALON-ZIPPEL-SCHWARTZ NULLSTELLENSATZ
LEMMA

Günter Rote

Abstract. We survey strengthenings and generalizations of the Schwartz-Zippel Lemma and Alon’s Combinatorial Nullstellensatz. Both lemmas guarantee the existence of (a certain number of) nonzero elements of a multivariate polynomial when the variables run independently through sufficiently large ranges.

1. Introduction
1.1. The Quantitative and the Existence Conclusion
1.2. Assumptions on the degrees \(d_i\)
1.3. Applications
1.4. Algebraic assumptions about the coefficients
1.5. Comparison of the assumptions
1.6. Tightness
1.7. Existence conclusions in the literature
1.8. Quantitative conclusions in the literature
1.9. Comparison of the assumptions
2. Remarks
2.1. Precursor results
3. Proof of Lemma Q
4. Largest total degree does not imply the Quantitative Conclusion
5. Proof of Lemma X by division by a linear factor
6. Weaker assumptions
6.1. Weaker assumptions for the Quantitative Conclusion
6.2. Weaker assumptions for the Existence Conclusion
7. Proof by Trimming
9. Weaker quantitative conclusions
9.1. Hypergraph model
10. What’s in a name?
References
Appendix A. Proof of Lemma X via the coefficient formula
Appendix B. Weaker algebraic assumptions, Condition (D)
Appendix C. OLD STUFF
1. Introduction

1.1. The Quantitative and the Existence Conclusion. Let \( f \in K[x_1, \ldots, x_n] \) be a polynomial in \( n \) variables over a field or integral domain \( K \), and let \( S_1, \ldots, S_n \) be subsets of \( K \). We want to make statements about the nonzeros of \( f(x_1, \ldots, x_n) \) when the variables \( x_i \) run independently over the sets \( S_i \), under the assumption that these sets are sufficiently large, compared to certain parameters \( d_1, \ldots, d_n \) that are related to the degrees of the terms in \( f \). We may then derive a mere conclusion about the existence of a nonzero or a stronger statement about the number of nonzeros:

**The Quantitative Conclusion.** If \( |S_i| > d_i \) for all \( i = 1, \ldots, n \), then the number of values \((x_1, \ldots, x_n) \in S_1 \times S_2 \times \cdots \times S_n \) such that \( f(x_1, \ldots, x_n) \neq 0 \) is at least

\[
(1) \quad (|S_1| - d_1) \cdot (|S_2| - d_2) \cdots (|S_n| - d_n)
\]

\[
(2) \quad = |S_1 \times S_2 \times \cdots \times S_n| \cdot (1 - \frac{d_1}{|S_1|}) (1 - \frac{d_2}{|S_2|}) \cdots (1 - \frac{d_n}{|S_n|}).
\]

In the term \((2)\), the factor on the right can be interpreted as a probability of getting a nonzero.

Since the product of the terms \(|S_i| - d_i\) is positive, an immediate consequence is

**The Existence Conclusion.** If \( |S_i| > d_i \) for all \( i = 1, \ldots, n \) then there exists a tuple of values \((x_1, \ldots, x_n) \in S_1 \times S_2 \times \cdots \times S_n \) such that \( f(x_1, \ldots, x_n) \neq 0 \).

1.2. Assumptions on the degrees \( d_i \). These conclusions hold under a variety of different assumptions about the parameters \( d_1, \ldots, d_n \).

To describe these parameters, we recall a few standard definitions. A **monomial** is a product \( x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \) of powers of variables \( x_i \) (not including a coefficient from \( K \)). The degree of the monomial in the variable \( x_i \) is the exponent \( a_i \), and the **total degree** is the sum \( a_1 + \cdots + a_n \) of these exponents.

The **monomials of a polynomial** \( f \) are the monomials that have nonzero coefficients when the polynomial is written out in expanded form as a linear combination of monomials.

The \((\text{partial})\) degree of a polynomial \( f \) in the variable \( x_i \) (or the degree of \( x_i \) in \( f \)) is the largest exponent \( a_i \) for which \( x_i^{a_i} \) appears as a factor of a monomial of \( f \). The total degree of a polynomial is the largest total degree of any of its monomials. This is what is usually called the degree of the polynomial, without further qualification.

A monomial of \( f \) is **maximal** if it does not divide another monomial of \( f \), see Figure \([\text{X}]\).

**Lemma X** (Generalized Combinatorial Nullstellensatz, Tao and Vu 2006 [19 Exercise 9.1.4, p. 332], Lasoń 2010 [10, Theorem 2]). If \( x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \) is a maximal monomial of \( f \), then the Existence Conclusion holds.

The **lexicographically largest** monomial \( x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \) of \( f \) is defined in the usual sense, see Figure \([\text{X}]\) if \( a_1 \) is the largest exponent of \( x_1 \) in all monomials of \( f \), \( a_2 \) is the largest exponent of \( x_2 \) in all monomials that contain \( x_1^{a_1} \) as a factor, \( a_3 \) is the largest exponent of \( x_3 \) in all monomials that contain \( x_1^{a_1} x_2^{a_2} \) as a factor, and so on. Of course, we may get a different lexicographically
largest monomial if we consider the variables in a different order. The results remain valid independently of the chosen order.

Figure 1. The forbidden monomials for the various assumptions are shown as grey regions, for \((d_1, d_2) = (4, 2)\).

**Lemma Q.** If the lexicographically largest monomial of \(f\) is \(x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n}\), then the Quantitative Conclusion holds.

1.3. **Applications.** Lemmas [Q] and [X] and their relatives (to be discussed shortly) have numerous important applications to combinatorics and algorithms. The results with the Quantitative Conclusion are the basis for many randomized algorithms, in particular for polynomial identity testing: Here one wants to check whether two polynomials are identical, or whether a given polynomial is identically zero. The polynomials are given by some algorithm that can evaluate them for specific values. Lemma [Q] provides a randomized test for this property, provided some a-priori bounds on the degree can be given.
When applying the results with the Existential Conclusion, in particular the Combinatorial Nullstellensatz (Corollary 1), a nonzero solution of the polynomial at hand represents some combinatorial object whose existence should be guaranteed.

The two application scenarios are interested in different ends of the probability spectrum. In randomized algorithms, the “success probability” should ideally be close to 1, but a reasonable probability that decays only polynomially to zero is good enough. Then, by choosing the parameters appropriately or by repeating the experiment, the success probability can be boosted to any desired security level. The precise probability bounds not so important in this context.

On the other hand, when it comes to questions of existence, the success of the argument comes down to whether the probability of having a non-zero is non-zero or not. Here it is important to know the smallest values $d_i$ for which the Existential Conclusion holds.

“The other are mentioned in Bishnoi Th 4.2 + 4.6”?

1.4. **Algebraic assumptions about the coefficients.** To a lesser extent, the various results in the literature differ in the assumption about the underlying ring of coefficients. We remark that all theorems that we state (with the exception of Lemma 14) hold when $K$ is an integral domain, i.e., a commutative ring without zero divisors. Section B mentions some weaker conditions on the ring.

1.5. **Comparison of the assumptions.** Figure 2 compares the strength of the various assumptions in these theorems, including some conditions that are defined in later sections.

The lexicographically largest condition of Lemma Q implies the maximality assumption of Lemma X but since the Quantitative Conclusion in Lemma Q is stronger than the Existence Conclusion in Lemma X, neither of the two results can be derived from the other, and there is no common generalization (see Section 4).

While maximality is not sufficient to imply the Quantitative Conclusion, there are some weaker quantitative conclusions that one can derive under the maximality assumption, see Section 9.

The assumptions in Lemmas X and Q for the Existence or the Quantitative Conclusion are not the weakest assumptions in terms of the monomials of $f$ that we are aware of. We mention some weakened assumptions in Section 6. They correspond to the two boxes in the top row of Figure 2.

1.6. **Tightness.** A simple family of polynomials shows that the bounds of Lemmas X and Q are tight: Select subsets $A_i \subset S_i$ of size $|A_i| = d_i$. Then the polynomial

$$
\prod_{i=1}^{n} \prod_{a \in A_i} (x_i - a)
$$

has degree $d_i$ in each variable $x_i$. The term $x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n}$ is simultaneously the lexicographically largest monomial and the unique maximal monomial, (and also the unique successively largest monomial in the sense of Lemma 9). It has $(|S_1| - d_1)(|S_2| - d_2) \ldots (|S_n| - d_n)$ zeros.
1.7. **Existence conclusions in the literature.** This is Alon’s original Combinatorial Nullstellensatz:

**Corollary 1** (Combinatorial Nullstellensatz, Alon 1999 [1, Theorem 1.2]). If $x_1^{d_1}x_2^{d_2}\ldots x_n^{d_n}$ is a monomial of largest total degree, then the Existence Conclusion holds.

Alon derives Corollary 1 from a companion result, [1, Theorem 1.1]. It states that, if the Existence Conclusion does not hold, and $f$ is zero on $S_1 \times S_2 \times \cdots \times S_n$, it can be represented in a certain way in the ideal generated by the polynomials $\prod_{a \in S_i}(x_i - a)$. This statement is analogous to Hilbert’s Nullstellensatz, and this justifies the name Combinatorial Nullstellensatz that Alon coined for these theorems. This algebraic statement can also be extended to the setting with weaker assumptions (Lemma X and Theorem 10, see [15, ...Section 7]), but we will focus on the Existence Conclusion.

1.8. **Quantitative conclusions in the literature.** The following bound follows by estimating the product $(1 - p_1)(1 - p_2)\ldots(1 - p_n)$ in [2] by the lower bound $1 - p_1 - p_2 - \cdots - p_n$. 

---

**Figure 2.** Relation between the assumptions on $d_1, \ldots, d_n$. The Quantitative and/or Existential Conclusion is indicated at the upper right corner of each box.
Corollary 2 (Schwartz 1979 [16, 17 Lemma 1]). Under the assumptions of Lemma Q, i.e., if the lexicographically largest monomial of \( f \) is \( x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n} \), the number of nonzeros is at least

\[
|S_1 \times S_2 \times \cdots \times S_n| \cdot \left(1 - \frac{d_1}{|S_1|} - \frac{d_2}{|S_2|} - \cdots - \frac{d_n}{|S_n|}\right).
\]

As a special case, when all sets \( S_i \) are equal, we get

Corollary 3 (The Schwartz-Zippel Lemma, Schwartz 1979 [16, 17, Corollary 1], see also [13, Theorem 7.2] or [19, Exercise 9.1.1, pp. 331–332]). If \( S_1 = S_2 = \cdots = S_n = S \) and the polynomial has total degree \( d \geq 0 \), then the number of nonzeros is at least

\[
|S|^n \cdot (1 - \frac{d}{|S|}).
\]

In other words, the probability of getting a zero of \( f \) if the variables \( x_i \) are uniformly and independently chosen from \( S \) is at most \( \frac{d}{|S|} \).

The probabilistic formulation with the upper bound \( d/|S| \) on the probability of getting a zero is the common statement of this lemma. The same holds for the following statements, but for comparison, we formulate all theorems in terms of the number of nonzeros.

The following statement looks at the degree of \( f \) in each variable \( x_i \):

Corollary 4 (Generalized DeMillo–Lipton–Zippel Theorem [3, Thm. 4.6], Knuth 1997, [8, Ex. 4.6.1–16, pp. 436]).

If \( d_i \) is the degree of variable \( x_i \) in \( f \), the Quantitative Conclusion holds.

Note that \( f \) does not have to contain the term \( x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n} \) in this case, but the powers occurring in the lexicographically largest monomial of \( f \) are at most \( d_i \).

As a special case, with a uniform bound on the degrees and all sets \( S_i \) equal, we get:

Corollary 5 (Zippel 1979 [20, Theorem 1, p. 221]). Suppose that \( f \) is not identically zero and the degree of each variable \( x_i \) in \( f \) is bounded by \( d \), and \( S_1 = S_2 = \cdots = S_n = S \). Then the number of nonzeros is at least

\[
(|S| - d)^n = |S|^n \cdot (1 - d/|S|)^n.
\]

The following statement puts a stronger assumption on \( d \):

Corollary 6 (DeMillo and Lipton 1978 [4, Inequality (1)]). If \( f \) has total degree \( d \geq 0 \) and \( S_1 = S_2 = \cdots = S_n = S = \{1, 2, \ldots, |S|\} \), then the number of nonzeros is at least

\[
|S|^n \cdot (1 - d/|S|)^n.
\]

Note that this has essentially the same assumptions as Corollary 3 (only the assumption about the set \( S \) is more specialized), but a weaker conclusion.

1.9. Comparison of the assumptions.

\footnote{see also Wikipedia, \url{http://en.wikipedia.org/wiki/Schwartz-Zippel_lemma} accessed 2022-01-16}
2. Remarks

The relation between the results in their published form is confusing. This is discussed [3, Section 4] and in the blog post.


https://anuragbishnoi.wordpress.com/2015/10/19/Alon-F"uredi,-Schwartz-Zippel,-DeMillo-Lipton-and-their-common-generalization/

As mentioned in Section 1.3, the precise bounds for the Qualitative Conclusion are of minor importance for the intended applications, and chose statement of the result may be the form that is most convenient to apply instead of the strongest form. Apparently Lemma [Q] was not written down simply because nobody cared to do so.


Precursor results. We mention two precursor results: In the first edition of Knuth's Art of Computer Programming, Vol. II, there is a weaker, qualitative version of the Quantitative Conclusion:

Corollary 7 (Knuth 1969 [7 Ex. 4.6.1–16, p. 379, solution on p. 54(2)). If \( f \) is not identically zero and \( S_1 = S_2 = \cdots = S_n = S = \{-N, -N + 1, \ldots, N - 1, N\} \), then the fraction of zeros of \( f \) in \( S_1 \times S_2 \times \cdots \times S_n \) goes to zero as \( N \rightarrow \infty \).

Øystein Ore, in 1922, already established the special case of the Schwartz-Zippel Lemma (Corollary [3]) when the variables \( x_i \) run over all elements of a finite field.

Corollary 8 (Ore 1922 [14, 11 Theorem 6.13]). If \( f \in \mathbb{F}_q[x_1, \ldots, x_n] \) is a polynomial of total degree \( d \geq 0 \) over a finite field \( \mathbb{F}_q \) and \( S_1 = S_2 = \cdots = S_n = \mathbb{F}_q \), then the number of nonzeros is at least \( (q^d - (q^d - 1)) \cdots (q^n - 1) \).

I have not been able to look are Ore’s paper myself, and I am citing it according to [11].

3. Proof of Lemma [Q]

Proof of Lemma [Q]. The proof is by induction on \( n \). The induction basis for \( n = 1 \) is the elementary fact that a degree-\( d \) polynomial has at most \( d \) zeros.

For \( n > 1 \), we write \( f \) in powers of \( x_1 \):

\[
(4) \quad f(x_1, \ldots, x_n) = \sum_{i=0}^{d_1} x_{1i}^{d_1} h_i(x_2, \ldots, x_n)
\]

The sum contains in particular the nonzero term \( x_{11}^{d_1} h_{d_1}(x_2, \ldots, x_n) \). By definition, \( x_{22}^{d_2} \ldots x_{n}^{d_n} \) is the lexicographically largest monomial of \( h_{d_1} \). By induction, the number \( N \) of tuples \( (x_2, \ldots, x_n) \in S_2 \times \cdots \times S_n \) for which \( h_{d_1}(x_2, \ldots, x_n) \neq 0 \) is at least

\[
N \geq (|S_2| - d_2) \cdots (|S_n| - d_n).
\]

\(^2\)In the second edition, this is on p. 418 and p. 620.
For a fixed \((x_2, \ldots, x_n)\) for which this case arises, \(f\) is a polynomial of degree \(d_1\) in \(x_1\). Therefore it has at most \(d_1\) zeros, and at least \(|S_1| - d_1\) nonzeros. Consequently, the number of nonzeros of \(f\) is at least

\[
(|S_1| - d_1)N \geq (|S_1| - d_1)(|S_2| - d_2) \cdots (|S_n| - d_n).
\]

**4. Largest total degree does not imply the Quantitative Conclusion**

We show that maximality (Lemma X) and not even largest total degree (Corollary 1) is sufficient to derive the Quantitative Conclusion. A counterexample is the polynomial

\[
f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2 - 1,
\]

describing an ellipse in the plane, and the sets \(S_1 = S_2 = \{-1, 0, 1\}\), see Figure 3. The monomial \(x_1x_2\) is a monomial of largest total degree, and the Quantitative Conclusion for \(d_1 = d_2 = 1\) would predict at least \((|S_1| - d_1)(|S_2| - d_2) = 4\) nonzeros on \(S_1 \times S_2\). However, there are only 3 nonzeros. (In fact, 3 is the smallest possible number of zeros in this case, for any polynomial for with \(x_1x_2\) as maximal monomial, because there must be a nonzero on each \(2 \times 2\) subrectangle of \(S_1 \times S_2\). SEE BELOW. also by knuth, Lemma 13)

![Figure 3. A quadratic bivariate polynomial with 6 zeros on a 3 x 3 grid](image)

**5. Proof of Lemma X by division by a linear factor**

We sketch the proof of Lasoni [10, Theorem 2], which extends the very simple proof of the original Combinatorial Nullstellensatz (Corollary 1) that was given by Michalek [12].

**Proof of Lemma X** We use induction on \(d_1 + \cdots + d_n\). The base case \(d_1 + \cdots + d_n = 0\) is obvious. Otherwise, assume w.l.o.g. that \(d_1 > 0\). Pick an element \(a \in S_1\) and divide \(f\) by \(x_1 - a\):

\[
f = q(x_1 - a) + r
\]

The remainder \(r\) is of degree 0 in \(x_1\), i.e., it is a function \(r(x_2, \ldots, x_n)\) and does not depend on \(x_1\). If \(r\) has a nonzero on \(S_2 \times \cdots \times S_n\), we obtain a nonzero of \(f\) by setting \(x_1 = a\). Suppose that \(r\) is zero on all of \(S_2 \times \cdots \times S_n\). Then we get a nonzero of \(f\) by finding a nonzero of \(q(x_1, x_2, \ldots, x_n)\) with \(x_1 \neq a\). The existence of such a nonzero in \((S_1 \setminus \{a\}) \times S_2 \times \cdots \times S_n\) is ensured by the inductive hypothesis: It is easy to check that \(x_1^{d_1-1}x_2^{d_2} \cdots x_n^{d_n}\) is indeed a maximal monomial of the quotient \(q\). 

Alternate Proofs:
Proof. PROOF 2.

"Reduction" (?) step, trimmed polynomial [15, Section 7]

Assume \(|S_i| = d_i + 1\).

The polynomial \(x_i^{d_i+1}\) and the degree-\(d_i\) polynomial \(x_i^{d_i+1} - \prod_{a \in S_i} (x_i - a)\) have the same values for all \(x \in S_i\). Hence, we may successively replace \(x_i^{d_i+1}\) by \(x_i^{d_i+1} - \prod_{a \in S_i} (x_i - a)\) and in this way, eliminate all powers of \(x_i\) higher than \(d_i\), without changing the value of \(f\) on \(S_1 \times S_2 \times \cdots \times S_n\).

In other words, we divide by

\[ h_i(x_i) := \prod_{a \in S_i} (x_i - a). \]

If we do this for all variables, we have a polynomial \(\tilde{f}\) (other names? \(F_n\) in Károlyi and Nagy) for which the degree in each variable \(x_i\) is \(d_i\). The maximal term \(x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n}\) is untouched by the reduction. It is therefore a lexicographically largest monomial, and we may apply Lemma Q (or even just Corollary 4).

In the appendix I give the proof as suggested which is the earliest proof of the generalized CN that I could trace.

The Combinatorial Nullstellensatz is a fundamental theorem that appears in a wide range of textbooks, and I have only skimmed the literature.

This proof, and all other proofs that I have seen proceed in two steps. They first reduce the polynomial \(f\) to a polynomial whose degree in each variable is less than \(|S_i|\), without changing the value of \(f\) on \(S_1 \times S_2 \times \cdots \times S_n\), and then apply an .. SZ-lemma-like. which is proved by induction on \(n\). The induction basis is that a univariate polynomial of degree \(d\) has at most \(d\) roots. At the very end, the core of this statement boils down to the division by a linear factor.

The proof of Michalet [12] puts the division by a linear factor at the very beginning.

By contrast, the above proof uses the the division by a linear factor. as its central step.

We see that the proof works for any integral domain: The only property from the coefficient ring that we need is that \(q(x_i - a)\) is nonzero whenever both factors are nonzero.

Bishnoi et al. [3] mention weaker constraints on the underlying algebraic structure. Condition (D): arbitrary ring, but none of the differences \(x - y\) for \(x, y \in S_i\) must be a zero divisor. For our proof, we need that those elements have inverses, for the Vandermonde determinant in the denominator to work.

6. WEAKER ASSUMPTIONS

There is a way in which the respective assumptions of Lemma X and Lemma Q can be weakened. The two variations of the assumptions were developed independently, but remarkably, they are similar in spirit. The assumptions become complicated and less natural, and they are motivated mainly by the fact that the original proof carries through with few changes.

6.1. WEAKER ASSUMPTIONS FOR THE QUANTITATIVE CONCLUSION. We define a more general notion than a lexicographically largest monomial, namely
a successively largest sequence $d_1, \ldots, d_n$ of exponents: We set $d_1$ to the degree of $x_1$ in $f$. Let $x^{e_1}$ be any power of $x_1$ that appears in a monomial of $f$, and let $f_2(x_2, \ldots, x_n)$ be the (nonzero) coefficient of $x^{e_1}$ in $f(x_1, \ldots, x_n)$. We then set $d_2$ to the degree of $x_2$ in $f_2$, and we choose any power $x_2^{j_2}$ of $x_j$ that appears in $f_2$ with a nonzero coefficient. We continue in this way: For $j = 3, \ldots, n$, assuming that $f_{j-1}$ and $e_{j-1}$ have been defined, let $f_j(x_j, \ldots, x_n)$ be the coefficient of $x_j^{e_{j-1}+1}$ in $f_{j-1}(x_{j-1}, \ldots, x_n)$, let $d_j$ be the degree of $x_j$ in $f_j$, and choose any power $x_j^{e_j}$ of $x_j$ that appears in $f_j$ with nonzero coefficient.

For example, if $f(x_1, x_2) = x_1^7 + x_1^6x_2^2 + x_1x_2 + x_2^6$, then $(d_1, d_2) = (7, 2)$ is a successively largest sequence, with the corresponding sequence $(e_1, e_2) = (1, 1)$. Another possibility is $(d_1, d_2) = (7, 6)$ with $(e_1, e_2) = (0, 6)$.

Note that $x_1^{d_1}x_2^{d_2} \ldots x_n^{d_n}$ is not necessarily a monomial of $f$, but $x_1^{e_1}x_2^{e_2} \ldots x_n^{e_n}$ is, with $e_j \leq d_j$. As with the lexicographically largest monomial, this notion also depends on the chosen order of the variables.

**Lemma 9** (Knuth, The Art of Computer Programming, [8] Answer to Ex. 4.6.1–16, pp. 675). For a successively largest sequence $d_1, \ldots, d_n$, the Quantitative Conclusion follows.

**Proof.** The proof of Lemma [Q] goes through with straightforward adaptations.

The proof is by induction on $n$. We write $f$ in powers of $x_1$ as in [4]:

$$f(x_1, \ldots, x_n) = \sum_{i=0}^{d_1} x_1^i h_i(x_2, \ldots, x_n)$$

By assumption, the sum contains in particular the nonzero term $x_1^{e_1} f_2(x_2, \ldots, x_n)$.

By definition, $(d_2, \ldots, d_n)$ is a successively largest sequence for $f_2$.

For a fixed tuple $(x_2, \ldots, x_n)$ with $f_2(x_2, \ldots, x_n) \neq 0$, $f$ is a nonzero polynomial of degree at most $d_1$ in $x_1$. In contrast to the case of Lemma [Q], the degree can be smaller than $d_1$, but the conclusion that $f$ has hat most $d_1$ zeros remains valid. The argument finishes in the same way as for Lemma [Q].

Knuth [8] Answer to Ex. 4.6.1–16] mentions further ideas of strengthening the bound, and points out the significance in the context of sparse polynomials.

### 6.2. Weaker assumptions for the Existence Conclusion.

**Theorem 10** (Schauz 2008 [15] Theorem 3.2(ii)). Assume $|S_i| > d_i \geq e_i$ for $i = 1, \ldots, n$. If $x_1^{e_1} \ldots x_n^{e_n}$ is a monomial in $f$ such that $f$ contains no other monomial $x_1^{e_1'} \ldots x_n^{e_n'}$ with $e_i' = e_i$ or $e_i' > d_i$ for each $i = 1, \ldots, n$, then the Existence Conclusion follows.

In the terminology of Schauz, $(e_1, \ldots, e_n)$ is called a “$(d_1, \ldots, d_n)$-leading multi-index”. The term $x_1^{d_1} \ldots x_n^{d_n}$ is not required to appear in $f$; thus, Theorem [10] may be stronger than Lemma [X]. For example if

$$f(x_1, x_2) = x_1^4x_2^8 + x_1x_2 + x_1^8x_2^2$$

We may take $(e_1, e_2) = (1, 1)$ and $(d_1, d_2) = (4, 2)$.
The forbidden exponent tuples are \(\{e_1, d_1 + 1, d_1 + 2, d_1 + 3, \ldots\} \times \{e_2, d_2 + 1, d_2 + 2, d_2 + 3, \ldots\}\).

... leaves \(e'_i\) unchanged or changes a point \(e'_i\) larger than \(d_i\) to a smaller one.

**Proposition 11.** If \((d_1, \ldots, d_n)\) is a successively largest degree sequence, as witnessed by \((e_1, \ldots, e_n)\), then the assumptions of Theorem 10 hold.

by the choice of \(d_1\), \(f\) contains no factors \(x^{e'_1}_1\) with \(e'_1 > d_1\) ....

Thus the terms that must be avoided are those with \(e'_1 = e_1\). These are the factors \(x^{e'_1}_1 f_2(x_2, \ldots, x_n)\).

... analogous to the relation between lexicographically largest and maximal monomials.

Proof 2 works! (PICTURE for a more generic 2d example). reduction does not change the coefficient of \(x^{e_1}_1 \ldots x^{e_n}_n\). \(\rightarrow\) nonzero polynomial where the degree in each variable is \(\leq d_i\). Quantitative.

By the multivariate Lagrange interpolation formula, such a polynomial (and hence the coefficient of \(x^{e_1}_1 \ldots x^{e_n}_n\) in the original polynomial \(f\)) is determined by the values of \(f\) on \(S_1 \times S_2 \times \cdots \times S_n\).

Actually, Schauz showed the stronger statement that the coefficient of \(x^{e_1}_1 \ldots x^{e_n}_n\) can be represented in terms of the values of \(f\) on \(S_1 \times S_2 \times \cdots \times S_n\), (generalizing the coefficient formula (13)).

In the applications of the Combinatorial Nullstellensatz or the Schwartz-Zippel Lemma and its relatives, the degree bounds on the polynomial \(f\) are derived ...

We are not aware of a “natural” application of Lemma \(X\) for which the classic Combinatorial Nullstellensatz (Corollary 1) would not suffice.

An application was given by Lasoñ [10, Theorem 4], gave an application of his Lemma \(X\) that goes beyond the classic Combinatorial Nullstellensatz (Corollary 1).

Moreover, However, for the combinatorial result about “lucky labelings” to which this theorem is applied, the polynomial \(f\) is a homogeneous polynomial. Thus all maximal monomials are also maximum-total-degree monomials.

The polynomial of [10] Theorem 4] is a small modification of a polynomial that can be obtained from a homogeneous polynomial \(h(x_1, \ldots, x_n)\) by replacing each variable \(x_i\) by some polynomial \(f_i(x_i)\).

We are not aware of any other

7. **Proof by Trimming**

Eliminating the powers \(x^{e_i}_i\) with \(e_i > |S_i|\).

Reduces a power \(x^{e'_i}_i\) with \(e_i > |S_i|\) to smaller powers of \(x_i\) while leaving all other factors \(x^{e'_j}_j\) unchanged.

Schauz: trimming does not change ....

8. **Stronger constraints: The Generalized Alon–Füredi Theorem**

The Generalized Alon–Füredi Theorem of Bishnoi, Clark, Potukuchi, and Schmitt [3] gives a precise bound on the minimum number of nonzeros when,
in addition to a bound \( d_i \) on the degree of each variable \( x_i \), the total degree \( d \) is specified. The bound is not explicit: It is formulated in terms of an optimization problem of minimizing the product of variables \( y_i \) under linear constraints:

**Theorem 12** (The Generalized Alon–Füredi Theorem, Bishnoi et al. [3]).

Let \( f \) be a polynomial of total degree \( d \), whose degree in each variable \( x_i \) is at most \( d_i \), where \( d_i < |S_i| \). Then \( f \) has at least \( N \) nonzeros on \( S_1 \times S_2 \times \cdots \times S_n \), where \( N \) is the optimum value of the following minimization problem:

\[
\text{minimize } y_1y_2\cdots y_n \\
\text{subject to } |S_i| - d_i \leq y_i \leq |S_i|, \text{ for } i = 1, \ldots, n \\
\sum_{i=1}^n y_i = |S_1| + \cdots + |S_n| - d
\]

\[\text{(6)}\]

**Proof.** Theorem 12 can be derived from Lemma Q. The optimization problem (6) can be interpreted as looking for a monomial \( x^{e_1}_1x^{e_2}_2\cdots x^{e_n}_n \) that is consistent with the assumptions of the Theorem, and for which Lemma Q gives the weakest bound.

To start the formal proof, note first that the optimum value \( N \) of (6) does not change if we turn (8) into an inequality:

\[\sum_{i=1}^n y_i \geq |S_1| + \cdots + |S_n| - d\]

\[\text{(8)}\]

This is easily seen as follows: Take a solution \((y_1, \ldots, y_n)\) satisfying (7) and (8). From the assumptions of the theorem, \( d \leq \sum_{i=1}^n d_i \). Therefore, as long as the inequality (8) is strict, one can always find a variable \( y_i \) that is not at its lower bound, i.e., \( y_i > |S_i| - d_i \). We can therefore reduce this variable, reducing the product \( y_1 \cdots y_n \).

Let \( x^{e_1}_1x^{e_2}_2\cdots x^{e_n}_n \) be the lexicographically largest monomial of \( f \). By the assumptions on \( f \), \( e_i \leq d_i \) and \( \sum_{i=1}^n e_i \leq d \). Hence, the quantities \( y_i := |S_i| - e_i \) satisfy the constraints (7)

\[|S_i| - d_i \leq y_i \leq |S_i|,\]
and the constraint (8′):
\[ \sum y_i \geq |S_1| + \cdots + |S_n| - \sum e_i \geq |S_1| + \cdots + |S_n| - d \]
By Lemma 10, the number of nonzeros is at least
\[ (|S_1| - e_1)(|S_2| - e_2) \cdots (|S_n| - e_n) = y_1 y_2 \cdots y_n, \]
which is at least the minimum value \( N \) of (6) under (7) and (8′).
\[ \square \]

Bishnoi et al. [3] proved Theorem 12 directly by induction on \( n \), and showed that the Generalized DeMillo–Lipton–Zippel Theorem (Corollary 4) can be derived from it.

The (original) Alon–Füredi Theorem [2, Theorem 5] differs from this theorem in an important point regarding the assumptions: The degrees \( d_i \) in the individual variables are not constrained, but on the other hand, it is assumed that \( f \) has at least one nonzero on \( S_1 \times S_2 \times \cdots \times S_n \).

Because of this additional assumption, the Alon–Füredi Theorem is not a trivial corollary of the Generalized Alon–Füredi Theorem, see [3, Sections 2.2–2.3]. In the constraints defining the bound \( N \), the lower bound in (7) is replaced by \( y_i \geq 1 \), and as a consequence, in contrast to Theorem 12, it is easy to solve the optimization problem: Starting from the lower bound \( y_1 = \cdots = y_n = 1 \), consider the variables \( y_i \) in order of decreasing sizes \( |S_i| \) and greedily enlarge each \( y_i \) value to its upper bound \( |S_i| \) until (8) is fulfilled.

We should mention that weaker condition on the ring bound is tight.

9. Weaker quantitative conclusions

Lemma 13. [Knuth, Mathematical Preliminaries Redux,Ex. MPR–114] If \( x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n} \) is a monomial of largest total degree over a field, then the \[ SOME WEAK FORM OF QUANTITATIVE CONCLUSION \] holds.
\[ \geq 1 + (|S_1| - (d_1 + 1)) + (|S_2| - (d_2 + 1)) + \cdots + (|S_n| - (d_n + 1)) \]

9.1. Hypergraph model. It is natural to associate an \( n \)-partite \( n \)-uniform hypergraph to the values of an \( n \)-variate polynomial over a grid \( S_1 \times \cdots \times S_n \): We put a hyperedge \((x_1, \ldots, x_n)\) whenever \( f(x_1, \ldots, x_n) = 0 \). The Existence Conclusion then says that the hypergraph contains no complete subhypergraph \( K^{(d_1 + 1, \ldots, d_n + 1)} \). We can ask what this statement alone (without regarding the algebraic background implies about the number of nonzeros in \( S_1 \times \cdots \times S_n \).
Erdős [5] Corollary, p. 188]. $r$-partite $r$-graph = $r$-uniform hypergraph with at least $n$ vertices in each color class, $n > n_0(r,l)$, with edge density
\[
> \frac{(3r)^r}{n^{1/r-1}}
\]
contains a complete $K^{(r)}(l, \ldots, l)$.

The $r$ of this statement is the number $n$ of variables in our setting, whereas the $n$ of this statement is the cardinality of $S$.

Translated from our setting, the Existence Conclusion says that the hypergraph corresponding to the zeros does not contain a complete $K^{(r)}(l, \ldots, l)$, with $l = d + 1$. We conclude that the density of zeros is bounded by
\[
(3n)^n / |S|^{1/(d+1)n-1}.
\]
if $|S|$ is big enough. This implies that the density of zeros goes indeed to 0 as $|S|$ is increased, but the convergence is very very slow.

KST: classic result of Kövari, Sós, Turán for $r = 2$.
\[
z(m,n;s,t) < (s - 1)^{1/t}(n - t + 1)m^{1-1/t} + (t - 1)m
\]

10. What’s in a name?

In the late 1970’s, the first randomized primality tests were discovered. Randomized algorithms were gaining popularity, and their strength was recognized. It may thus be no coincidence that various forms of the Schwartz-Zippel Lemma were discovered independently, as the topic was “in the air”.

The papers of Schwartz and Zippel were even presented at the same conference and published back to back in the proceedings volume.

For some reason, the name Schwartz-Zippel Lemma stuck, despite the accumulation of sibilant consonants, and despite the priority of DeMillo and Lipton. A blog post of Richard Lipton discusses this question and speculates about the reasons for this fact [6, 18] http://people.maths.ox.ac.uk/keevash/papers/turan-survey.pdf. We add to this discussion by speculating that the quirk with the capital letter in the middle of the family name might have caused some insecurity and uneasiness. The poor typsetting quality of the journal Information Processing Letters at the time may have further contributed to the fact that the paper was not sufficiently received. In the title of this note, we honor the tradition of omitting DeMillo and Lipton. As we saw, Lasoč’s generalization of Alon’s Combinatorial Nullstellensatz was predated by an exercise in a textbook, but he must be credited for bringing the explicit statement of Lemma $X$ to the published journal literature, and his name is also included for the rhyme.

References


Appendix A. Proof of Lemma \[ \text{IX} \] via the coefficient formula

This proof follows the hint of Tao and Vu \[19\] Exercise 9.1.4, p. 332\], see also Lasoń \[10\] Section 3\]. Essentially the same proof, for the case of the original Combinatorial Nullstellensatz (Corollary \[1\]), was given by Kouba \[9\] in 2009.

As an intermediate result, we get a formula \[13\] for the coefficient of $x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n}$ in terms of the values of $f$ on $S_1 \times S_2 \times \cdots \times S_n$ (the Coefficient Formula of Lasoń \[10\] Theorem 3\]).

We emphasize, that in contrast to other statements, the following proof does suppose that $F$ is a field.
We start with a preparatory lemma:

**Lemma 14.** Let $\mathbb{F}$ be a field. For a nonempty set $S \subseteq \mathbb{F}$, there is a function $g = g_S : S \to \mathbb{F}$ with the following property:

\[
\sum_{x \in S} g(x)x^k = 0, \text{ for } k = 0, 1, \ldots, |S| - 2
\]

\[
\sum_{x \in S} g(x)x^k = 1, \text{ for } k = |S| - 1
\]

**Proof.** The equations (9, 10) form a system of $|S|$ linear equations in the $|S|$ unknowns $u_j = g(a_j)$ for $a_j \in S = \{a_1, a_2, \ldots, a_{|S|}\}$. The coefficient matrix is a Vandermonde matrix, and hence the system has a unique solution.

The solutions $u_j$ can actually be obtained explicitly as the quotient of two Vandermonde determinants:

\[
u_j = g(a_j) = 1 / \prod_{k \neq j} (a_j - a_k) \quad \square \]

**Proof of Lemma 14.** It is no loss of generality to assume $|S_i| = d_i + 1$. Take the functions $g_{S_i}$ for $i = 1, \ldots, n$, and multiply them together:

\[
g(x_1, \ldots, x_n) := g_{S_1}(x_1)g_{S_2}(x_2) \cdots g_{S_n}(x_n)
\]

Following the suggestion of Tao and Vu [19, Exercise 9.1.4] further, we consider the quantity

\[
\tilde{F} := \sum_{x_1 \in S_1} \sum_{x_2 \in S_2} \cdots \sum_{x_n \in S_n} f(x_1, \ldots, x_n)\tilde{g}(x_1, \ldots, x_n)
\]

and show that $\tilde{F} \neq 0$. Let us see how this transformation affects the monomials $x_1^{a_1} \cdots x_n^{a_n}$ of $f$:

\[
\sum_{x_1 \in S_1} \sum_{x_2 \in S_2} \cdots \sum_{x_n \in S_n} x_1^{a_1} \cdots x_n^{a_n} g_{S_1}(x_1)g_{S_2}(x_2) \cdots g_{S_n}(x_n)
\]

\[
= \sum_{x_1 \in S_1} x_1^{a_1} g_{S_1}(x_1) \cdot \sum_{x_2 \in S_2} x_2^{a_2} g_{S_2}(x_2) \cdots \sum_{x_n \in S_n} x_n^{a_n} g_{S_n}(x_n)
\]

This expression vanishes whenever $a_i < d_i$ for some $i$, by [9]. The only monomial of $f$ that is not annihilated in this way is the maximal term $x_1^{d_1}x_2^{d_2} \cdots x_n^{d_n}$. For this term, the sum (14) becomes 1, by (10). Therefore $\tilde{F}$ as given by (13) is the coefficient of $x_1^{d_1}x_2^{d_2} \cdots x_n^{d_n}$ in $f$. By assumption, $\tilde{F} \neq 0$. Therefore, by (13), there must be an $(x_1, x_2, \ldots, x_n) \in S_1 \times S_2 \times \cdots \times S_n$ with $f(x_1, \ldots, x_n) \neq 0$.

The hint of Tao and Vu [19, Exercise 9.1.4] actually suggests to prove a more general version of Lemma [14].

**Lemma 15.** For a set $S$ with $|S| > d$, there is a function $g = g_{S,d} : S \to \mathbb{R}$ with the following property:

\[
\sum_{x \in S} g(x)x^k = \begin{cases} 0, & \text{for } k = 0, 1, \ldots, d - 1 \\ 1, & \text{for } k = d \end{cases}
\]
This can be derived by applying Lemma 14 to an arbitrary subset $S' \subseteq S$ of size $|S'| = d + 1$ and setting $g_{S,d}(x) = 0$ for $x \notin S'$. We have instead chosen to simplify the proof by assuming $|S| = d + 1$.

Since we are constructing some sort of interpolating function $g$, which depends on solving a system of equations, this proof depends on $\mathbb{F}$ being a field (or at least, a ring in which all nonzero differences $a - a'$ for $a, a' \in S_i$ are units).

Tao and Vu [19, Exercise 9.1.4] formulate their exercise “for a field whose characteristic is 0 or greater than $\max d_i$.” I don’t see how the characteristic of the coefficient field comes into play.

Bishnoi et al. [3] mention weaker constraints on the underlying algebraic structure. Condition (D): arbitrary ring, but none of the differences $x - y$ for $x, y \in S_i$ must be a zero divisor. For our proof, we need that those elements have inverses, for the Vandermonde determinant in the denominator to work.

**APPENDIX B. WEAKER ALGEBRAIC ASSUMPTIONS, CONDITION (D)**

none of the differences $a - a'$ for $a, a' \in S_i$, $a - a'$ is a zero divisor.

**APPENDIX C. OLD STUFF**

plays a crucial role in complexity

Dana Moshkovitz:

The Schwartz-Zippel lemma is remarkably important to theoretical computer science: it is used for polynomial identity testing (e.g., to test whether a graph has a perfect matching), it establishes the distance property of the Reed-Muller code (the code that consists of truth tables of multivariate low degree polynomials), and it is used in proofs of hardness vs. randomness tradeoffs, as well as in algebraic constructions of Probabilistically Checkable Proofs. The lemma has an inductive proof. In this note we present a different, more direct, proof.

An Alternative Proof of The Schwartz-Zippel Lemma, Dana Moshkovitz, Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 96 (2010). only over a finite field! $\leq d/|\mathbb{F}|$.


FROM THE INTERNET: The Schwartz-Zippel lemma (see [51], pg. 165) is a probability result which is commonly used as the basis of a random algorithm for verifying proposed polynomial identities. Suppose we are given a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, and we suspect $f$ is identically zero, but the act of deterministically expanding and simplifying $f$ takes exponential time (in the number of variables). Therefore, we are interested in a method for randomly verifying $f$ in polynomial-time (in the number of variables), and determining whether $f$ is identically zero within some user-specified, error bound. Note that, since a Nullstellensatz certificate is identically one, subtracting one from the certificate creates an identically zero polynomial. The Schwartz-Zippel lemma is as follows:

in fact it implies Schwartz-Zippel. This is exercise 9.1.1 in Tao and Vu’s Additive Combinatorics, right after they introduce the combinatorial Nullstellensatz. [G.R.: THIS IS NOT TRUE. The writer did not read to the end of the exercise.] They remark that it is ”particularly useful for obtaining lower bounds on the size of restricted sum sets and similar objects.” – Qiaochu Yuan Oct 23 2010 at 19:19

FOLLOWS PROBABLY FROM THE SAME PROOF.

**Corollary 16** (NON-Zippel[20]). Suppose that the degree of variable $x_i$ in $f$ is $d_i$. Then, with this interpretation of $d_i$, the number of nonzeros is at least as given by (2).

GY. KÁROLYI AND Z. L. NAGY: https://arxiv.org/pdf/1211.6484.pdf to express this coefficient we apply the following effective version of the Combinatorial Nullstellensatz [1] observed independently by Lasoń [19] and by Karasev and Petrov [15]. A sketch of the proof is included for the sake of completeness.

Karasev and Petrov, Theorem 4: in terms of the derivatives $\phi'(\cdot)$ in the denominator.


From Anurag Bishnoi’s remark on Lipton’s block.

PhD thesis of Daniel Edwin Erickson from 1974, Counting Zeros of Polynomials over Finite Fields, a Theorem 5.1. with $s = 0$


.... A more natural application would be desirable.

readily applied

common consequences

3 http://mathoverflow.net/questions/43317/how-to-recognise-that-the-polynomial-method-might-work